Local Endpoint-Constrained Residual Calibration
A three-location temperature test

Adrian Dunkley

Abstract

We train separate temperature reconstruction networks at sample points in Jamaica, Barbados and Puerto Rico, then fit an explicit residual equation using a separate calibration year. Each local model has independent fitted weights. The equation uses endpoint temperature change and local diurnal and seasonal phase as empirical predictors. It preserves supplied endpoint temperatures but does not impose an atmospheric energy balance. Training uses 2019–2020, calibration 2021, selection 2022 and evaluation 2023–2024. Calibration improves Jamaica's validation error yet worsens test root-mean-square error from 0.1992 to 0.2035 degrees Celsius. Barbados changes negligibly and Puerto Rico rejects the correction on validation. Cross-location coefficient transfer offers no general improvement. The result supports local testing and rejection rules, not a newly discovered physical law.

1 Introduction

Local reconstruction errors may contain repeatable structure, but fitting a correction does not establish that it will improve a later year. This empirical retrospective supplement tests an explicit residual equation after training a separate neural model at each location. The question is whether calibration-year phase and endpoint relationships improve held-out reconstruction, and whether fitted coefficients transfer between locally trained models. The contribution is an auditable test of this hypothesis, including rejection and adverse outcomes.

2 Data and model-selection methods

Local Endpoint-Constrained Residual Calibration names the workflow, not a claim of methodological novelty. We fit three distinct 273-parameter networks with two hidden layers of 16 and 8 units. Inputs are two known six-hour endpoint temperatures, fractional hour and cyclic hour/day coordinates. Three random seeds and two checkpoints per seed are retained for each location. The training optimiser, architecture and source data match the accompanying temporal methods study; no fitted weights are shared across locations. Each island is represented by one gridded point, not an island-wide observational sample.

NASA POWER hourly temperatures identify MERRA-2 reanalysis as their source. The experiment reconstructs five intermediate hours when both six-hour endpoints are already known. Targets are not station observations, and no operational forecast is tested. Root-mean-square error (RMSE), mean absolute error (MAE) and signed bias are evaluated in degrees Celsius; adjacent-hour rate error is evaluated in degrees Celsius per hour. There is no classification accuracy percentage. All locations have 14,610 withheld intermediate test records across 2023–2024.

Neural weights and normalisation use only 2019–2020 at the relevant location. The 2021 residuals fit correction coefficients. The 2022 RMSE selects the uncorrected checkpoint and representative seed, then selects correction regularisation for that fixed model. The 2023–2024 records fit nothing and select nothing. These years had been inspected in earlier development, so this extension is exploratory rather than confirmatory.

3 Methods: residual equation and adoption rule

Let $\alpha \in [0, 1]$ be the fractional position between endpoint temperatures T_0, T_6, and $F = \alpha(1-\alpha)$. The local neural estimate is

$$\hat{T}_{\mathrm{N}} = (1 - \alpha)T_0 + \alpha T_6 + F f_\theta(x).$$

The correction candidate is

$$\begin{aligned}\hat{T}_{\mathrm{C}} = \hat{T}_{\mathrm{N}} + F\big[&\beta_0 + \beta_1 \Delta T + \beta_2 \sin\phi + \beta_3 \cos\phi \\ &+ \beta_4 \sin\psi + \beta_5 \cos\psi + \beta_6 \Delta T \cos\phi\big],\end{aligned}$$

$$\Delta T = T_6 - T_0, \qquad \phi = \frac{2\pi}{24}(h_{\mathrm{UTC}} + \ell/15), \qquad \psi = \frac{2\pi d}{365.25}.$$

Here ℓ is longitude in degrees east, d is day of year and h_{UTC} is hour in Coordinated Universal Time. Mean solar time is approximated using longitude; equation-of-time effects are omitted. The phase variables and endpoint trend give contextual proxies for diurnal reconstruction errors. They are not radiation, advection or turbulent-flux observations.

The seven coefficients minimise squared 2021 residual error plus a ridge penalty. Penalties 0.01, 1, 100 and 10,000 are compared using 2022 only. Features are not standardised for this small regression; their stated units and exact coefficients are retained. Coefficients multiplying ΔT are dimensionless; the remaining coefficients have temperature units. The factor F preserves both endpoints exactly for every coefficient value.

A validation rejection rule retains the neural estimate when the correction is no better on 2022. The recorded development history states that this rule was clarified after inspecting validation diagnostics but before test metrics were read; this timing is not externally registered. Every forced-correction candidate remains available separately from the adopted policy. The basis was specified explicitly: this is residual regression, not automated symbolic discovery of equation structure.

4 Results and coefficient-transfer tests

Location	Linear	Local calendar	Local neural	Calibration policy
Jamaica	0.56547	0.20986	0.19919	0.20346
Barbados	0.12536	0.08085	0.08102	0.08103
Puerto Rico	0.66914	0.23733	0.21302	0.21302

Table 1: 2023–2024 RMSE in degrees Celsius. Local calendar corrections use 2019–2021, matching the neural-plus-calibration system's total fitting/calibration years.

The Jamaica correction was favoured by validation but worsened test performance. Its increase in mean squared error is 0.00172 squared degrees Celsius; 7-day and 30-day circular-block sensitivity intervals remain positive. Barbados has a negligible worsening, with the 30-day interval spanning zero. Puerto Rico's chosen correction was worse on validation, so the policy retains the original local network exactly. A simple calendar remains marginally better at Barbados. No additional correction is justified by the aggregate test results.

The seven-term equation can be interpreted as a testable hypothesis: local residuals may follow a repeatable diurnal pattern interacting with interval temperature change. The experiment does not support the stronger assertion that this fitted relationship generalises across years and locations. Applying Jamaica's coefficients to Barbados increases RMSE from 0.08102 to 0.08553 degrees Celsius, and applying them to Puerto Rico increases it from 0.21302 to 0.21854. The transfer matrix evaluates fitted candidate coefficients, including the rejected Puerto Rico

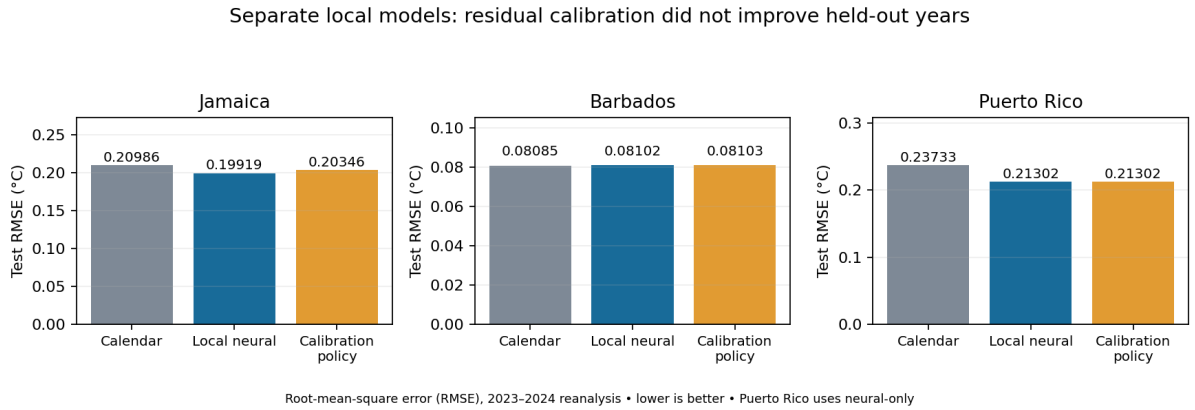


Figure 1: Separate local models and the validation-based calibration policy. Panel scales differ. The test outcome is not used to retroactively select a more favourable correction.

candidate. This transfers only coefficients onto each destination's local network, not a complete model to an unseen island. The accompanying leave-one-island-out study addresses that different question.

First-half and second-half 2021 fits are retained as sensitivity diagnostics. Different seasonal coverage confounds coefficient comparisons, so disagreement cannot establish physical instability. Circular blocks preserve limited short-range dependence but do not eliminate uncertainty from product choice, site selection or interannual variability.

5 Discussion: simple corrections and physical interpretation

A post-hoc comparison fits constant bias and one-parameter endpoint-preserving bias on 2021, then applies the same validation rejection rule. Both are rejected for Jamaica and Barbados. Puerto Rico accepts both on validation, but neither improves test RMSE. The comparator expansion occurred after initial test inspection; coefficients still use only calibration data. Constant bias does not preserve endpoints, while the endpoint-bias candidate does. All candidates and decisions are retained.

A neural residual does not uniquely identify omitted physics. Reanalysis smoothing, model bias, measurement representation and missing covariates can produce similar error patterns. A physical air-temperature equation would need compatible advection, flux divergence, storage and boundary/source terms; an assumed surface budget cannot simply be substituted. No computational efficiency gain is measured: the correction adds operations and does not replace the neural model. This study does not identify a new climate theory or a universal correction formula.

Physics-guided weather downscaling and neural temporal interpolation already exist[1, 2, 3]. Their existence prevents treating a regional application or a penalty term as novelty by itself. The present evidence is narrower: explicit local calibration can fail despite validation improvement, and coefficient transfer must be tested rather than assumed.

6 Limitations and future work

Acquire independent hourly station targets and compatible meteorological forcing. Test whether residual phase patterns recur across multiple locations within each island, while fitting no information from the evaluation years. Compare a physically justified closure against matched empirical and no-physics controls. Use archived issue-time forecasts for operational questions and retain reanalysis reconstruction as a separate experiment. Reserve genuinely untouched

years or stations for confirmation, with sufficient events to estimate uncertainty. A journal-level contribution would require those stronger comparisons and a clearly demonstrated advance; acceptance is not implied by naming this workflow.

Source data, all 18 neural checkpoints, 36 ridge candidates, selected coefficients, year-level metrics, candidate-versus-policy distinctions and review records are preserved. The earlier models and reports are unchanged.

7 Supplementary coefficient record

Table 2 gives the fitted candidate coefficients in the stated basis and units. Puerto Rico's candidate is not adopted by the validation policy. Rounded displayed values are explanatory; full precision remains in the coefficient files.

Coefficient	Jamaica	Barbados	Puerto Rico
β_0	0.0308611	-0.0000548	0.0012374
β_1	-0.0003454	0.0002751	0.0069130
β_2	-0.1518353	0.0001045	0.0020213
β_3	-0.1255580	-0.0006122	-0.0009099
β_4	0.0077531	-0.0007114	-0.0000756
β_5	-0.1084704	-0.0004432	-0.0014123
β_6	0.0191564	-0.0001670	-0.0023034

Table 2: Empirical residual coefficients fitted on 2021. They are not coefficients of an identified atmospheric conservation law.

8 Conclusion

Separate local networks permit location-specific calibration, but the tested residual equation does not deliver consistent improvements. Jamaica's validation gain reverses on the test years, Barbados changes negligibly and Puerto Rico retains the uncorrected model. The evidence supports explicit adoption checks and further hypothesis testing rather than a universal correction or discovered climate law.

9 Data and code availability

The source data, neural checkpoints, correction candidates and diagnostic records are retained privately. No public archive of the analysis materials is currently provided. External methodological references below should not be interpreted as links to this study's private model files.

References

[1] Luo et al. (2025). Physics-Guided Learning of Meteorological Dynamics for Weather Downscaling and Forecasting. ACM SIGKDD. https://doi.org/10.1145/3711896.3737081.

[2] Wang et al. (2025). Neural Network Based on Dynamic Collaboration of Flows for Temporal Downscaling. Remote Sensing 17, 1434. https://doi.org/10.3390/rs17081434.

[3] Serifi et al. (2021). Spatio-Temporal Downscaling of Climate Data Using Convolutional and Error-Predicting Neural Networks. Frontiers in Climate 3, 656479. https://doi.org/10.3389/fclim.2021.656479.